

# USC Marshall

School of Business

## DSO 528: DATA WAREHOUSING, BUSINESS INTELLIGENCE AND DATA MINING

**Online - Spring 2020 – Saturday 9:00-11:50 A.M. – First Draft**

**Instructor:** Dr. Arif Ansari  
**Office:** BRI 401 R (Bridge Hall 401R)  
**Office Hours:** Zoom Appointment and in person Appointment  
**Phone:** (213) 821-5521  
**Email:** [aansari@marshall.usc.edu](mailto:aansari@marshall.usc.edu),  
**Emergency Contact number:** 213-740-0172

### COURSE DESCRIPTION

We live in age of Big Data, in this era, Analytics, Data-driven Decision Making and Data-Driven Strategies are the main source of getting Competitive Advantage in the turbulent Business World. **The focus of the course is to make current students and future business professionals Analytically savvy.** This course will enable students to leverage data, analyze data, build quantitative models, interpret the results and make data-driven decisions. The important 'big picture' goal of this course is to use Critical Thinking & Creative Thinking to build analytical models using holistic view of the company, data sources of the company, relevant external big data and to make decision under uncertainty, a most sought-after skill in all business professions.

The topics covered are: Data Mining Models (Prediction Models, Classification Models, Clustering Models and Association Models), Introduction to Big Data, Different types of Analytics, Introduction to data-warehousing, multidimensional database, on-line analytical processing and business intelligence.

### COURSE OBJECTIVES

Upon successful completion of this course, students will be able to:

1. Understand Data Warehousing (DW), Big Data (BD) and Business Intelligence
2. Develop data mining skills to monetize data - Perform basic data mining analysis and understand analyses performed by others (e.g. consultants) - applying an integrated approach to understanding and analyzing significant business problems, which can be complex, messy, unstructured, and beyond formulaic analysis
3. Develop desktop level data mining using SAS JMP software and be familiar with industry level "SAS-Enterprise Miner" tool (if time permits)..
4. Develop quantitative reasoning to support written arguments - Using critical and analytical thinking to identify viable solutions that can create short-term and long-term value for organizations.
5. Learn to select appropriate analytical techniques for analyzing the data - devising creative, sustainable, and achievable strategies and solutions that allow organizations to take advantage of opportunities that create value for its stakeholders.
6. Apply Descriptive, Predictive, Prescriptive and Creative Analytics used in Business.

## COURSE MATERIALS

The following items will be necessary for completion of cases, team project and successful completion of the course.

### 1. Class notes and Class Videos

Class notes for this class will be available on blackboard. You should familiarize yourself with these notes before they are covered in class. In addition, I will post class videos for your preparation and I expect you to watch them before Class.

If you have any questions or need assistance with the Blackboard Course Pages, please contact the Marshall HelpDesk at 213-740-3000 or HelpDesk@marshall.usc.edu

### 2. Text Books:

#### Problems:

1. **Big Data Analytics is a wide area to be covered in a single textbook.**
2. **In general, the text books are either business oriented (less emphasis on techniques and methods) or subject oriented (less emphasis on real world problems).**
3. **The textbooks tend to solve standard problems not the ill-defined problems that are common in Big Data Analytics world.**

#### Solution:

1. **I will cover most of the topics in my PowerPoint slides. In addition, I will give you access to Teradata University which has a lot of material on Data warehousing, Databases, BI etc. I will post information from IBM to cover Big Data material.**
2. **The cases will give you hands on experience in building models and using tools.**

Suggested Text Book: The Book I would recommend for the class from JMP tool perspective is

**Ron Klimberg and B. D. McCullough (2013), Fundamentals of Predictive Analytics with JMP, SAS Institute. ISBN-13: 978-1612904252 (Recommended)**

**This book shows you how to use JMP for building models, it is a relatively low cost book but it does not have the textbook type structure to it.**

### Additional suggested Books for this class

- The first book is a standard book for Data Mining, the book talks about the various techniques and it is written from computer science perspective. **(Recommended)**

Data Mining: Concepts and Techniques, Second Edition by Jiawei Han and Micheline Kamber, Morgan Kaufmann Publishers, ISBN 13: 978-1-55860-901-3, ISBN-10: 1-55860-901-6, website: [www.mkp.com](http://www.mkp.com)

**Note: The book is written from a Computer Science and it will help you to understand the data mining techniques but it does not have real world business application – Buy the book if you want to understand Data Mining Algorithms. My PowerPoint slides will cover the data mining topics but not in depth.**

- The second book is from SAS – The world’s leading Data mining software company. This book introduces you to industry level Data mining software – SAS Enterprise Miner. **(Recommended)**

Data Mining Using SAS Enterprise Miner – A Case Study Approach, Second Edition . ISBN 1-59047-190-3, SAS publishing

website: [www.sas.com](http://www.sas.com)

Note: The book helps you to get hands on experience in real-world tool and teaches how to solve well-defined problems.

- The third book is from SAS. This book introduces you to industry level Data mining software – SAS Enterprise Miner. **(Recommended)**

Data Mining Using SAS Enterprise Miner by **Randall Matignon**, John Wiley and Sons publishing

website: [www.sasenterpriseminer.com](http://www.sasenterpriseminer.com)

Note: The book helps you to learn SAS model building methods and get hands on experience in real-world tool and teaches how to solve well-defined problems.

#### **Additional Books of Interest (If you want to concentrate on Data Warehousing)**

- Building the Data Warehouse 3rd Edition, W.H. Inmon, Wiley, ISBN 0-471-08130-2
- Data Warehouse: Practical Advice from the experts, Joyce Bischoff and Ted Alexander, Prentice hall, ISBN 0-13-577370-9
- Recommended: Data Warehousing: using the Wal-mart model. *Paul Westerman, Morgan Kauffman publishers.*

#### **Online Resources**

- **Sign up with Teradata University Network**

Teradata University Network ([www.TeradataUniversityNetwork.com](http://www.TeradataUniversityNetwork.com)) is a **free** learning portal designed to help faculty to teach, learn, and connect with others in the fields of data warehousing, DSS/Business Intelligence, and database.

Teradata University offers web-based courses and related web sites on data warehousing, DSS/BI and database. They have a library of Teradata white papers. Students can become Teradata certified. We will use their material and software in the class particularly for the Business Intelligence and Data warehousing part of the class.

Students register for [www.TeradataStudentNetwork.com](http://www.TeradataStudentNetwork.com) and login using the current password: \*\*\*\* (I will provide the password if I get Teradata University Approval)

#### **EXPECTED LEARNING OUTCOMES**

- You will get to know the Big Data Analytics Domain

- You will be able to frame and solve Business Analytical problems
- You will be capable of building predictive models
- You will get hands on data mining skills to monetize data
- **6 in-depth real-world cases on Online Marketing, Search Engine Marketing, Loan Default, Churn, Ensemble Modeling and Data Mart**
- You will become familiar with SAS JMP software (If time permits I will introduce industry level “SAS-Enterprise Miner” tool)
- You will learn to interpret and communicate the key business insights obtained from model building
- You will become familiar with Data Warehousing
- You be able to Envision, Manage and Lead Analytical Projects (Entry Level)

## COURSE STRUCTURE

- **70% of the class will be focused on Data Mining**
- **10% on Business Intelligence tools**
- **20% on Data warehousing**

## Overview:

Big Data has made available a lot of Structured and Unstructured data along with it new Paradigms for Monetizing data. Many of these new paradigms are ill-defined problems and businesses are struggling to develop methods and models to leverage the “Big Data Opportunity.” Most entry level Analytics professionals are “man-in-the-middle” between data and business requirements; they don’t know how to fully utilize all the available data and are not fully aware of the domain expertise needed to build efficient business savvy models. This course is for students who want to be “Standing on the Shoulders of Giants (Big Data Analytics)” and have great vision on the data side and on the business side, understand Big Data - its potential and drawbacks, Statistics - its usefulness and limitations, data mining - its usefulness and limitations, Business needs and available opportunity. In short this class is about how to monetize Data in customer facing applications using critical thinking and creative thinking.

Companies have huge amount of data in their data warehouse and have access to Big Data through 3<sup>rd</sup> party APIs. Companies want to leverage data for decision making by building “Data Driven Decision Making Models” and they want to monetize big data using data mining (DM) and Business Intelligence.

The access to social, demographic, transactional, click-stream, web usage etc., data has made companies “data rich” and now they want new ways to monetize data as wells as enhance the traditional predictive models using Enriched data. For example, Fortune 500 companies such as American Express, Wells Fargo and Wal-mart have accumulated a great deal of data from their day to day business now they want to monetize the data by providing value to customers and sell their products and services through Omni channels in an efficient manner.

In the Big Data Analytics space what are critical are information, knowledge, insight and monetization. Some of the questions are: what is the utility of the data? How can one use data in managing customer relationship and empowering employees? How can one uncover patterns and relationships hidden in databases? How can one creatively find ways to monetize data through analytical models? How can one enhance the performance of existing models?

In summary, managers need to understand the strategic values of their company's information assets, be capable of building analytical models to monetize data, understand the models built by third party companies, be able to extract insights from the models and be able to visualize data and insights.

## COURSE GOALS:

By Topic,

- 1) In Data Mining you will develop in-depth skill set to do desktop Data Mining and learn the industry level Data Mining tool.
- 2) In Data Warehousing/Big Data (DW/BD) part you will learn, why companies need DW/BD, advantages of DW/BD and how to create a DW schema that an executive will understand, I will not teach the hands-on programming for DW part, DW programming part is made available through Teradata student network and you can learn on your own.
- 3) In Business Intelligence you will learn what current BI can do, how to develop the requirements of a BI system for a company. I will not teach the hands-on programming part, programming part is made available through Teradata student network (and Tableau) and you can learn on your own.
- 4) You will learn how the 3 parts are interconnected and integrated to form the basis of corporate knowledge system. How to leverage them to convert your company to near real-time corporation. How to monetize data.
- 5) **Identify, Conceive, Formulate and Solve Predictive Analytics Problems.**

### Structure of lectures:

DSO 528 will be organized in a way that includes some combination of the following: lectures, case-based class discussion and computer work.

This class is designed in such a way that only limited mathematical and statistical (Descriptive Statistics, Hypothesis testing and Regression) background is required. I will give a brief review on the above-mentioned topics. Learning and understanding underlying DW/BD concepts, studying cases, applying DM/BD ideas and methods to business data, and communicating ideas and solutions will be our main theme. Technical details of selected DM methods will be discussed. Students are expected to use Data Mining software for various cases in class.

### *Synchronous Sessions (Saturday Class Sessions)*

In order to earn full participation points, students must actively participate in all synchronous sessions via computer or laptop, with a webcam and headset/speakers. Students are expected to be visually present and to ask thought-provoking questions, offer relevant comments, and answer questions from faculty in a clear and concise manner.

As outlined in the student handbook, there are specific expectations of an online student. When attending online live sessions, present and act appropriate as if you were in a physical classroom. Examples of inappropriate actions include, but are not limited to:

- Inappropriate dress, such as revealing clothing, costumes, pajamas, bathrobes or swim attire
- Engaging in a simultaneous activity (e.g., using a telephone, reading a book, knitting)
- Interacting with persons who are not part of the class
- Leaving frequently or not being on camera for extended periods of time
- Having other persons or pets in view of the camera
- Actively engaging in child care
- Behaving in an overtly inattentive manner (looking distracted, not participating)

- Attending class in a non-quiet area where activities of other people or pets are potential distractions
- Engaging in inappropriate tone and language with instructors or classmates during discussions, postings, instant messaging or email

#### *Discussion Forums*

In order to earn full written discussion forum points, students must write thoughtful, robust posts as well as substantive responses to their peers.

#### *Groups*

In addition to individual assignments, students will select their own groups of to work on team assignments. Unless otherwise specified, they will work within these course teams to complete all group or team assignments during this semester.

#### *Communication*

Students are expected to notify instructor(s) for prior approval if they are unable to attend a synchronous session or to complete a written discussion, assignment, or exam by the posted due date.

#### **Asynchronous Activities**

All readings, videos, and assignments can be found in blackboard.

### **GRADING**

Your final course grade will represent how you performed in the class relative to other students. Your grade will not be based on a mandated target, but on your performance. Historically, the average grade for a graduate elective class at USC Marshall is about a 3.5 average. Your grade will be based on the following components (refer below\_

Assignments	% of Overall Grade
Class Participation (including written submission)	15%
Cases	25%
Team Project	20%
Mid-Term Exam	20%
Final Exam	20%
TOTAL	100%

### **CLASS PARTICIPATION**

Class participation is an extremely important part of the learning experience in this course as the richness of the learning experience will be largely dependent upon the degree of preparation by *all* students prior to each class session. In addition to class discussion, you will submit your answer in written format via email. Refer to Appendix III

#### **Cases**

We will analyze six cases during the semester. The cases will be evaluated and will be counted towards the case points. The Cases can be done in groups of 4 or 5 or 6 students. You will turn in 5

out of 6 cases. Since the Group size is large, **each of you will add your independent Case Summary in addition to Group Case Submission.** You will do a peer evaluation for each case, if needed.

### TEAM PROJECT

I strongly believe the students learn the most by doing a datamining project. The Group will consist of 4, 5 or 6 students. Learning to work in teams is essential and to get different perspective and will greatly enhance your learning. Since the Group size is large, **each of you will add your independent Project Summary and your sample model in addition to Group Case Submission.** You will be getting your own dataset (need to be approved by Professor) and submit the project in multiple phases. Refer to the Project Guideline posted in Blackboard.

### MIDTERM (or Quizzes) and FINAL EXAM

Midterm exam will take place at the beginning of class for approximately one hour and 45 minutes. You may bring two sheets (four pages) containing formulas, definitions etc., to the midterm except solved problems and solved multiple choice questions. For the final, you may bring four sheets (eight pages) containing formulas, definitions etc., except solved problems and solved multiple choice questions. No make-ups of mid-term or the final will be given. You will receive a grade of zero for each missed exam unless you have a written excuse from your doctor or the professor. In case of emergency or approved absence, the professor may decide to give a make-up exam or redistribute the points.

### **Information on Cases**

**Case 1 – Prof. Ansari’s Smart Partyware – “Find the high propensity customers for Celebrating American Arts product”**

**Learn Key Concepts like Perfect Marketing, Decision Tree, Beating Computer Models, Understanding Data Mining Metrics, KPIs and Lift**

The Smart Partyware Company’s business model is direct-to-consumer marketing. Over the years they have gained dedicated upscale customers and currently have 500,000 members in their database.

In the direct-marketing industry, the response rate is measured as a percentage of customers who buy the directly mailed product. Smart Partyware’s historical response rate for direct mail to selected members is approximately 10% – far above the industry average. SPW was using RFM (Recency-Frequency-Monetary) analysis to target customers. Smart Partyware wants to increase the response rate well beyond the 10% rate.

SPW designs new party ware for every campaign, gives a new name to its party ware, and broadly classifies the party ware under one of its many party themes. Most of the designs cut across many themes but are classified into a particular category based on the main design theme in the party ware. The recent product to be marketed is Celebrating American Arts. It has famous American art works printed in the party ware and the objective is to find high propensity customers for the current marketing campaign. In this case the students will be using Decision Tree Model.



## **Case 2 – Prof. Ansari’s Search Engine Marketing for Smart Partyware – “Selecting the right set of keywords for search engine marketing campaign.”**

### **Learn Key Concepts like Search Engine Marketing, Dash Boards, Clustering and Campaign Management**

The Smart Partyware (SPW) Company’s business model is direct-to-consumer marketing. Over the years they have gained dedicated upscale customers and currently have 500,000 members in their database.

Applichem has signed a Memorandum of Understanding (MOU) with SPW. They will acquire 10 percent of SPW for an undisclosed sum and have an option to buy up to a total of 49% in the following year at current valuation determined by independent evaluators.

John Runner one of the founders of SPW has a vested interest – he wanted to increase the revenue and profit of SPW so that the valuation of SPW in a year will be high and Applichem will have to pay more for the shares of SPW. John and other executives’ contracts with SPW allowed them to sell up to 25% of their shares as part of the deal with Applichem. John Runner was sure his prodigy Vijay would be able to do his magic once again and would be able to increase revenue and profit.

Vijay knew he had fully leveraged the power of data mining; increasing the efficiency of the algorithms would not increase the revenue and profit by 50%. His first approach was to buy a potential member list from data brokers to increase the number of members at SPW. This approach was not successful; the additional revenue from new members was not substantial. In fact, the profit from new members was negligible after taking into account the amount of money paid for the data acquisition and the cost of phone-based marketing to enroll them as new members. The second approach was revamping the site and doing Search Engine Marketing (SEM).

SPW signed up with Google AdSense and created an account with Google. Based on “Partyware” search wording Google AdSense gave a list of nearly 800 keywords and phrases that people normally search, along with the level of competition, the number of local monthly searches, and the approximate cost per click (CPC). The total amount spent per month on the “partyware” keyword was approximately \$250,000. SPW agreed to allocate \$20,000 for ad budget the first month, and based on the success or failure the next month’s budget would be decided. Based on the keyword bidding SPW wants to sign-up as many visitors to its website as members and increase its membership base.

The challenge of this case is to find the right cluster(s) of keywords for SEM campaign so that many prospects will visit the website and join as members.

## **Case 3 – SAS’s “Home Equity”**

### **Learn Key Concepts like Logistic Regression, Profiler, Odds, KPIs, Lift and Monetization**

The consumer credit department of a bank wants to automate the decision-making process for approval of home equity lines of credit. To do this, they will follow the recommendations of the Equal Credit Opportunity Act to create an empirically derived and statistically sound credit scoring model. The model will be based on data collected from applicants granted credit through the process of loan underwriting. The model will be built from predictive modeling tools, but the created model must be sufficiently interpretable so as to provide a reason for any adverse actions (rejections).

The HMEQ data set contains baseline and loan performance information for 1000 recent home equity loans. The target (BAD) is a binary variable that indicates if an applicant eventually defaulted or was seriously delinquent. This adverse outcome occurred in approximately 10% of the cases. The challenge of this case is to predict the BAD loans so the consumer credit department will be able to prevent default and make appropriate decision on the home equity line of credit.

#### **Case 4 - New Case Churn using Neural Network**

##### **Learn Key Concepts like Neural Network, Transformation and Profiler**

Customer retention is a challenge in the ultracompetitive mobile phone industry. A mobile phone (service provider) company is studying factors related to customer churn, a term used for customers who have moved to another service provider.

##### **The Task**

The company would like to build a model to predict which customers are most likely to move their service

Currently there are 1 million customer accounts, this month the expected churn rate is 15%. The monthly revenue is \$50 per customer per month. It will cost \$10 discount per month to keep customer who might churn away from churning. The Current Revenue is \$50 million. If all the people who want to churn leave the company then the revenue will be \$42.751 Million.

Assume the company is willing to review up to 15% accounts that they think will churn and give \$10 proactive discount of \$10 (promotion).

It is important to correctly identify the potential churner among the million customers.

##### **Data-Mining Business Models**

The mobile company believes the best method to use is Neural Network as they will be able to defend the model based on non-linear modelling.

Help Build a model to identify the churners and maximize the new net revenue per month (refer to excel sheet for calculations).

## **Case 5 – Prof. Ansari’s Smart Partyware – Big Data Approach – Data Blending and Ensemble Modeling**

### **Learn Key Concepts like Big Data, Data Blending, Ensemble Modeling and Leveraging Multiple DM methods**

The Smart Partyware Company’s business model is direct-to-consumer marketing. Over the years they have gained dedicated upscale customers and currently have 500,000 members in their database.

In the direct-marketing industry, the response rate is measured as a percentage of customers who buy the directly mailed product. Smart Partyware’s historical response rate for direct mail to selected members is approximately 10% – far above the industry average. SPW was using RFM (Recency-Frequency-Monetary) analysis to target customers. Smart Partyware wants to increase the response rate well beyond the 10% rate.

SPW designs new party ware for every campaign, gives a new name to its party ware, and broadly classifies the party ware under one of its many party themes. Most of the designs cut across many themes but are classified into a particular category based on the main design theme in the party ware. The recent product to be marketed is Celebrating American Arts. It has famous American art works printed in the party ware and the objective is to find high propensity customers for the current marketing campaign. In this case the students will be using the standard business analytics methods like, Decision Tree, Logistics Regression and Neural Network.

## **Case 6 – Star Schema – Model Trojan Dental School Business Intelligence Engine.**

### **Learn Key Concepts like Data Mart, Star Schema, KPIs and BI**

Trojan Dental School is involved in continuing education program for Dentists. This is revenue generating area for the Dental School and they want to increase the revenue and efficiency in targeting the potential candidates for continuing education.

The Dental School has Alumni data and can buy the list of Dentists in Southern California area from a data-broker. You are hired by Trojan Dental School to create Business Intelligence Engine that will enable the dental school to become the premier continuing education program for Dentists in Southern California region, to increase revenue from its Continuing Education Program, to estimate its Customer Life time value, to improve its Customer Service, to Create actionable dashboards for its Administrators and faculty and make their continuing education program an agile organization.

<b>STATEMENT OF ACADEMIC CONDUCT AND SUPPORT SYSTEMS</b>
--

USC seeks to maintain an optimal learning environment. Students are expected to submit original work. They have an obligation both to protect their own work from misuse and to avoid using another’s work as their own. All students are expected to understand and abide by the principles of

academic honesty outlined in the University Student Conduct Code (see University Governance, Section 11.00) of SCampus ([www.usc.edu/scampus](http://www.usc.edu/scampus) or <http://scampus.usc.edu>). The recommended sanctions for academic integrity violations can be found in Appendix A of the Student Conduct Code.

### **Students with Disabilities:**

USC is committed to making reasonable accommodations to assist individuals with disabilities in reaching their academic potential. If you have a disability which may impact your performance, attendance, or grades in this course and require accommodations, you must first register with the Office of Disability Services and Programs ([www.usc.edu/disability](http://www.usc.edu/disability)). DSP provides certification for students with disabilities and helps arrange the relevant accommodations. Any student requesting academic accommodations based on a disability is required to register with Disability Services and Programs (DSP) each semester. A letter of verification for approved accommodations can be obtained from DSP. Please be sure the letter is delivered to me (or to your TA) as early in the semester as possible. DSP is located in GFS (Grace Ford Salvatori Hall) 120 and is open 8:30 a.m.–5:00 p.m., Monday through Friday. The phone number for DSP is (213) 740-0776. Email: [ability@usc.edu](mailto:ability@usc.edu).

### **Support Systems:**

*Student Counseling Services (SCS) - (213) 740-7711 – 24/7 on call*

Free and confidential mental health treatment for students, including short-term psychotherapy, group counseling, stress fitness workshops, and crisis intervention. <https://engemannshc.usc.edu/counseling/>

*National Suicide Prevention Lifeline - 1-800-273-8255*

Provides free and confidential emotional support to people in suicidal crisis or emotional distress 24 hours a day, 7 days a week. <http://www.suicidepreventionlifeline.org>

*Relationship & Sexual Violence Prevention Services (RSVP) - (213) 740-4900 - 24/7 on call*

Free and confidential therapy services, workshops, and training for situations related to gender-based harm. <https://engemannshc.usc.edu/rsvp/>

*Sexual Assault Resource Center*

For more information about how to get help or help a survivor, rights, reporting options, and additional resources, visit the website: <http://sarc.usc.edu/>

*Office of Equity and Diversity (OED)/Title IX compliance – (213) 740-5086*

Works with faculty, staff, visitors, applicants, and students around issues of protected class. <https://equity.usc.edu/>

*Bias Assessment Response and Support*

Incidents of bias, hate crimes and microaggressions need to be reported allowing for appropriate investigation and response. <https://studentaffairs.usc.edu/bias-assessment-response-support/>

*Student Support & Advocacy – (213) 821-4710*

Assists students and families in resolving complex issues adversely affecting their success as a student EX: personal, financial, and academic. <https://studentaffairs.usc.edu/ssa/>

*Diversity at USC – <https://diversity.usc.edu/>*

Tab for Events, Programs and Training, Task Force (including representatives for each school), Chronology, Participate, Resources for Students

### *Emergency Preparations*

In case of an emergency if travel to campus is not feasible, the USC Emergency Information web site (<http://emergency.usc.edu/>) will provide relevant information, such as the electronic means the instructors might use to conduct their lectures through a combination of USC's Blackboard learning management system ([blackboard.usc.edu](http://blackboard.usc.edu)), teleconferencing, and other technologies.

<b>COURSE OUTLINE AND ASSIGNMENTS</b>
---------------------------------------

**Tentative Schedule:**

---

- I. The course will start with Data Mining. The Data Mining part of the class will be quantitative and the following topics will be covered in it.

1. Standard Data Mining techniques:
  - a. Classification
  - b. Clustering
  - c. Association
  - d. Visual Data mining

Using various appropriate techniques,

- i) Bayesian Estimation
- ii) Neural Networks
- iii) Decision Tree
- iv) Nearest Neighbor
- v) Other techniques like Ensemble Modeling

2. Statistical Model Building using Logistic Regression.  
Depending on the project other topics may be covered.

- II. The second part of the course will be Data Warehousing. You will be introduced to Data Warehousing from business perspective, how to create Data Warehouse Architecture.

- III. In addition to the above, if time permits we will discuss the following,

**Advanced Data Mining Concepts**

## Schedule of class – Subject to change

TUN - Teradata University Network, SAS - Enterprise Miner Text book

JM - Data Mining textbook by Jiawei Han and Micheline Kamber

Bring Your Laptop to class with JMP software

Date		Topic	Reading from textbooks	Reading Class notes	Due/Other
1/18/20	1	Introduction to Big Data Analytics Introduction to Classification - Distance Based Algorithms/Critical Thinking Exercise	JM 1-26, JM 36-40, JM 359-362, JM372-375 JM285-290, JM 310-318, JM 347-350, SAS 39-67	Dr. Ansari Notes	Bring in Laptop with JMP for all lectures
1/25/20	2	Descriptive Statistics/Classification Methods Tuple Discussion/ – Decision Tree Based Methods/Case 1		Dr. Ansari Notes	
2/01/20	3	Decision Tree Based Methods Evaluation/ Project Discussion	JM 291-306, SAS 19-36	Dr. Ansari Notes	Turn in your Individual Case1 Group List Due
2/08/20	4	Decision Tree Strategies/Review of Hypothesis Testing	SAS 39-67, SAS 67-81	Dr. Ansari Notes	Turn in your Case1
2/15/20	5	Search Engine Marketing Clustering and Association/ Google Analytics and Adwords	JM 227-234, SAS 91-104 JM 358-359, JM 327-336	Dr. Ansari Notes	Your Project Topic
2/22/20	6	Complete Case2/ Logistic Regression/Review of Hypothesis Testing	JM 384-414, JM 227-234	Dr. Ansari Notes	Decide your Project and Dataset Turn in your Individual Case2
2/29/20	7	Logistic Regression Continued/ Case 3 Bayesian/ Nearest Neighbor/	JM 358-359, JM 327-336	Dr. Ansari Notes	Turn in your Case2
3/07/20	8	Bayesian/ Nearest Neighbor/ Review	JM 384-414, JM 227-234, SAS 91-104, SAS 105-109	Dr. Ansari Notes	Submit your Project and Data set Submit it in class Turn in your Individual Case3
3/14/20	9	Midterm /Neural Network		Dr. Ansari Notes	Turn in your Case3
3/21/20	10	Spring Break			
3/28/20	11	Neural Network/Case4/ Association		Dr. Ansari Notes	Turn in your Preliminary report Turn in your Individual Case4
4/04/20	12	Business Intelligence/Data Enrichment/ Case5/Ensemble Model		Dr. Ansari Notes	Turn in your Case4
4/11/20	13	Lecture DW1: Data Warehousing(I): Strategic View Lecture DW2: A Tactical View Lecture DW3: Dimensionally Designed DW	JM 105-114, JM 127-134 and TUN relevant information JM 114-123 and TUN relevant information JM 123-126 and TUN relevant information	Dr. Ansari Notes	Turn in your Individual Case5 Oral Presentation of Project (if Time Permits)
4/18/20		Lecture DW4: OLAP and Business Intelligence Lecture Reporting/Presentation/Review	JM 135-137, JM 144-152 and TUN relevant information	Dr. Ansari Notes	Turn in your Case5 Oral Presentation of Project and Turn in your Final Report
4/25/20	14	Presentation/Review	JM 135-137, JM 144-152 and TUN relevant information	Dr. Ansari Notes	Turn in your Case6 Oral Presentation of Project and Turn in your Final Report
4/27/20	15	Presentation/Review			
5/02/20 Or 5/09/20		Final Exam – 2-4 p.m.			

## Appendix I. MARSHALL GRADUATE PROGRAMS LEARNING GOALS

### How DSO 528 Contributes to Marshall Graduate Program Learning Goals

Marshall Graduate Program Learning Goals	DSO 528 Objectives that support this goal	Assessment Method*
<b><i>Learning Goal #1: Develop Personal Strengths.</i></b> <b>Our graduates will develop a global and entrepreneurial mindset, lead with integrity, purpose and ethical perspective, and draw value from diversity and inclusion.</b>		
1.1 Possess personal integrity and a commitment to an organization's purpose and core values.		
1.2 Expand awareness with a global and entrepreneurial mindset, drawing value from diversity and inclusion.		
1.3 Exhibit awareness of ethical dimensions and professional standards in decision making.		
<b><i>Learning Goal #2: Gain Knowledge and Skills.</i></b> <b>Our graduates will develop a deep understanding of the key functions of business enterprises and will be able to identify and take advantage of opportunities in a complex, uncertain and dynamic business environment using critical and analytical thinking skills.</b>		
2.1 Gain knowledge of the key functions of business enterprises.	1	Cases
2.2 Acquire advanced skills to understand and analyze significant business opportunities, which can be complex, uncertain and dynamic.	2	Cases
2.3 Use critical and analytical thinking to identify viable options that can create short-term and long-term value for organizations and their stakeholders.	5,6	Project
<b><i>Learning Goal #3: Motivate and Build High Performing Teams.</i></b> <b>Our graduates will achieve results by fostering collaboration, communication and adaptability on individual, team, and organization levels.</b>		
3.1 Motivate and work with colleagues, partners, and other stakeholders to achieve organizational purposes.		
3.2 Help build and sustain high-performing teams by infusing teams with a variety of perspectives, talents, and skills and aligning individual success with team success and with overall organizational success.	3,4	Project
3.3 Foster collaboration, communication and adaptability in helping organizations excel in a changing business landscape.		





## Appendix II

### SAMPLE PEER EVALUATION FORM

Please identify your team and team members for the Cases and Project that you worked on. Then rate all your team members, *including yourself*, based on the **contributions** of each team member for the selected assignment according to the criteria listed below. On a scale of 0 – 2 with 0 indicating does not meet expectations, 1 meets expectations and 2 exceeds expectations, rate each person on each of the five criteria. Lastly, add up the points for each person with the maximum number of points for each person being 10. In the box below, describe the exact contributions of each team member, including yourself.

Team Members/ Assessment Criteria of Team Contributions	Team Member 1	Team Member 2	Team Member 3	Team Member 4	Team Member 5	Yourself
1. Role Performance						
2. Assists Team Members						
3. Listening and Discussing						
4. Research and Information Sharing						
5. Time Management						
<b>Total</b>						

Contribution details:

### Appendix III

Class participation is an extremely important part of the learning experience in this course as the richness of the learning experience will be largely dependent upon the degree of preparation by *all* students prior to each class session.

A course that incorporates the frequent use of case analyses to illustrate the practical application of concepts and practices requires the student to diligently and thoroughly prepare cases and actively offer the results of the analyses and conclusions derived as well as recommendations during each class session. My expectation and that of your classmates are that you are prepared for *all* classes and will actively participate in and meaningfully contribute to class discussions.

In-class participation is also a critical part of this course's learning experience. Cold calling may take place to encourage active participation and to gain multiple perspectives and points of view, thus lending itself to the richness of the learning experience. In-class participation grading will be based on students' demonstrated willingness to participate and the quality of the comments expressed, rather than quantity. While some students are far more comfortable than others with class participation, *all* students should make an effort to contribute meaningfully.

Students will offer their opinions in group settings many times in their careers; thus, class participation serves to prepare students for this business experience.

The evaluating of in-class participation is based on the following:

- *Relevance* – Does the comment or question meaningfully bear on the subject at hand? Irrelevant or inappropriate comments can detract from the learning experience.
- *Responsiveness* – Does the comment or question connect to what someone else has said?
- *Analysis* – Is the reasoning employed consistent and logical? Has data from course materials, personal experience, or general knowledge been employed to support the assertions/findings?
- *Value* – Does the contribution further the understanding of the issues at hand?
- *Clarity* – Is the comment concise and understandable?

During class sessions, I frequently assume the role of a facilitator to encourage a discussion that includes perspectives from a variety of viewpoints and, secondly, to help pull together prevailing analyses and recommendations. The direction and quality of a discussion is the *collective responsibility of the class*.

**You will turn in your Question and Answer via email to me for Credit.**

For each in-class session two (2) points will be awarded to a student for relevant and meaningful participation, one (1) point for modest contributions to the class and zero (0) points for no participation or absence.

To underscore the importance of participation, 5 percent of the course grade is allocated to class participation.

## Class Participation—Behavioral Anchor Rating Scale:

### Excellent Performance

- Initiates information relative to topics discussed
- Accurately exhibits knowledge of assignment content
- Clarifies points that others may not understand
- Shares personal experiences or opinions related to topic
- Offers relevant / succinct input to class
- Actively participates in class exercises
- Demonstrates ability to apply, analyze, evaluate & synthesize course material.
- Demonstrates willingness to attempt to answer unpopular questions
- Builds on other students' contributions

### Average Performance

- Participates in group discussions when asked
- Demonstrates knowledge of course material
- Offers clear, concise, “good” information on class assignments
- Offers input, but tends to reiterate the intuitive
- Attends class regularly

### Unacceptable Performance

- Fails to participate even when directly asked
- Gives no input to discussions
- Does not demonstrate knowledge of the readings
- Shows up to class: does nothing
- Distracts group / class
- Irrelevant discussion

## Appendix IV

### MIDPOINT COURSE EVALUATION QUESTIONS

*Marshall Faculty are encouraged to give students midpoint course evaluations to gauge student concerns and adjust the course early on. Student feedback is for instructor use only and not a part of the formal performance review process. Instructors are encouraged to review the comments and discuss in the following class session.*

In order to continuously improve the effectiveness of our class, could you please take a few moments to answer the following questions:

1. How well do the course objectives support your general business knowledge and personal career goals?
2. What have you liked about this course so far?
3. Do you have any suggestions for improving the course experience?